

TMS136

Föreläsning 9

Olika skattningsmetoder

- Givet ett stickprov finns det en mängd olika metoder för att göra *punktskattningar (point estimates)* av fördelnings-/populationsparametrar
- En punktskattning är ett tal som är en (förhoppningsvis kvalificerad) gissning av ett parametervärde
- Vi ska kika på två olika metoder:
- *Momentmetoden* som handlar om att skatta teoretiska moment med stickprovsmoment
- *Maximum likelihood* som handlar om att hitta vilka parametervärden som "troligast ligger bakom" stickprovet

Momentmetoden

- Vi kallar $E(X^m)$ *m-temomentet* för X
- $m = 1$ ger förstamomentet eller väntevärdet
- $m = 2$ ger andramomentet
- Varians är alltså "andramomentet minus första momentet i kvadrat"
- Givet stickprovet $X_1, \dots, X_n$ kan vi skatta $E(X^m)$ med $\frac{1}{n} \sum_{i=1}^n X_i^m$ (*m-te stickprovsmomentet*)

Momentmetoden

- Låt $X_1, \dots, X_n$ vara stickprov från en fördelning med okända parametrar $\theta_1, \dots, \theta_k$.
- *Momentskattarna* $\hat{\theta}_1, \dots, \hat{\theta}_k$ fås genom att ersätta de k första populationsmomenten $E(X), \dots, E(X^k)$ med de k första stickprovsmomenten $\frac{1}{n} \sum_{i=1}^n X_i, \dots, \frac{1}{n} \sum_{i=1}^n X_i^k$
- Alltså; Om parametern θ_i kan uttryckas som en funktion $\theta_i = f\left(E(X), \dots, E(X^k)\right)$ fås momentskattaren ur $\hat{\theta}_i = f\left(\frac{1}{n} \sum_{i=1}^n X_i, \dots, \frac{1}{n} \sum_{i=1}^n X_i^k\right)$ för samma funktion f

Exempel

- Antag att stickprovet $X_1, \dots, X_n$ kommer från en exponentialfördelning med okänd parameter λ och vi vill skatta λ med dess momentskattare
- Vi vet att $E(X_i) = \frac{1}{\lambda}$. Vi skattar $E(X_i)$ med $\bar{X}$ och får momentskattaren $\hat{\lambda} = \frac{1}{\bar{X}}$ för λ
- Om vi observerat $x_1, \dots, x_n$ blir $\frac{1}{\bar{x}}$ vår skattning av λ

Exempel

- Antag att stickprovet $X_1, \dots, X_n$ kommer från en normalfördelning med okända parametrar μ och σ^2 och vi vill skatta μ och σ^2 med dess momentskattare
- Vi vet att $E(X_i) = \mu$ och $E(X_i^2) - (E(X_i))^2 = \sigma^2$ så vi får momentskattarna

$$\hat{\mu} = \bar{X}$$

och

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$$

Maximum Likelihood

- Låt $x_1, \dots, x_n$ vara ett observerat stickprov från en fördelning med täthet/massfunktion f och okända parametrar $\theta_1, \dots, \theta_k$. *ML-skattarna* $\hat{\theta}_1, \dots, \hat{\theta}_k$ fås genom att maximera *likelihood-funktionen*

$$L(\theta_1, \dots, \theta_k) = \prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_k)$$

med avseende på $\theta_1, \dots, \theta_k$

Maximum Likelihood

- Intuitionen är att ML ger de parametervärden som mest troligt har genererat stickprovet
- I många fall är det behändigare att arbeta med *log-likelihood-funktionen*

$$l(\theta_1, \dots, \theta_k) = \ln L(\theta_1, \dots, \theta_k)$$

- Maximering av l med avseende på $\theta_1, \dots, \theta_k$ ger samma skattningar som maximering av L med avseende på $\theta_1, \dots, \theta_k$

Exempel

- Antag att vi observerat $x_1, \dots, x_n$ från en Bernoulli-fördelning och vill skatta parametern p
- Då har vi att $f(x; p) = p^x(1 - p)^{1-x}, x = 0,1$
- Likelihood-funktionen blir

$$L(p) = \prod_{i=1}^n p^{x_i}(1 - p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}$$

Exempel

- Att maximera L är bökigt, så vi väljer att arbeta med

$$l(p) = \ln p \sum_{i=1}^n x_i + \ln(1 - p)(n - \sum_{i=1}^n x_i)$$

- Derivering med avseende på p ger

$$\frac{dl}{dp} = \frac{1}{p} \sum_{i=1}^n x_i - \frac{1}{1-p} \left(n - \sum_{i=1}^n x_i \right)$$

Exempel

- Sätter vi derivatan lika med noll får vi

$$p = \frac{1}{n} \sum_{i=1}^n x_i$$

Så ML-skattaren för p är alltså $\hat{p} = \bar{X}$ och om vi observerat $x_1, \dots, x_n$ blir $\bar{x}$ vår skattning av p

Exempel

- Antag att vi observerat $x_1, \dots, x_n$ från en normalfördelning och vill "ML-skatta" parameterarna μ och σ^2
- Vi har att likelihood-funktionen är

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2} \right\} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \end{aligned}$$

Exempel

- "Log-likelihooden" blir

$$l(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

- Derivering m.a.p. μ ger $\mu = \bar{x}$ så ML-skattaren för μ är

$$\hat{\mu} = \bar{X}$$

Exempel

- Derivering av l m.a.p. σ^2 ger

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

så ML-skattaren för σ^2 är

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$$

Observation

- Vi ser att Momentskattaren och ML-skattaren för variansen i en normalfördelning "är samma"
- Den är inte vvr, ty (verifiera likheterna)

$$E(\widehat{\sigma^2}) = \frac{1}{n} \sum_{i=1}^n E(X_i^2) - E((\bar{X})^2) = \sigma^2 \left(1 - \frac{1}{n}\right)$$

- Man kan alltså förvänta sig en underskattad varians om man använder denna skattaren för små stickprov. Å andra sidan vet vi att S^2 är vvr så vi kanske hellre använder den för små stickprov

Egenskaper hos ML-skattare

- Under generella villkor gäller att ML-skattare är
- Asymptotiskt vvr, alltså bias avtar då stickprovsstorleken växer
- De har liten varians
- De är approximativt normalfördelade

Invarians hos ML-skattare

- Om $\hat{\theta}_1, \dots, \hat{\theta}_k$ är ML-skattarna för parametrarna $\theta_1, \dots, \theta_k$ gäller för någon funktion h att $h(\hat{\theta}_1, \dots, \hat{\theta}_k)$ är ML-skattare för $h(\theta_1, \dots, \theta_k)$

- Exempel: Vi såg att ML-skattaren för variansen i en normalfördelning är

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$$

- Invariansegenskapen ger att $\sqrt{\widehat{\sigma^2}}$ är ML-skattaren för standardavvikelsen i samma normalfördelning. Här är alltså $h(x) = \sqrt{x}$

Nackdelar med ML

- Det är inte alltid så lätt att maximera likelihood-funktionen..
- Om exempelvis X är likformigt fördelad på intervallet $[0, a]$ har vi att tätheten för X är $1/a$ på intervallet och noll annars.
- För ett stickprov av storlek n blir likelihood-funktionen (som vi använder vid skattning av a)

$$L(a) = \frac{1}{a^n}$$

Nackdelar med ML

- Om vi deriverar likelihood-funktionen får vi

$$\frac{dL}{da} = -na^{n-1}$$

som ju är negativ för $a > 0$. Så vi kan inte hitta ML-skattaren för a med våra "vanliga" analysmetoder

- Men eftersom L är avtagande i a har vi att L maximeras om vi, givet ett observerat stickprov $x_1, \dots, x_n$, sätter $a = \max(x_i)$
- ML-skattaren för a är alltså $\hat{a} = \max(X_i)$
- Detta är ju någorlunda rimligt eftersom a aldrig kan vara mindre än den största stickprovsobservationen (eller hur?)

Egenskaper hos ML-skattaren för övre gränsen i likformig fördelning

- Vi har alltså att ML-skattaren för a i en likformig fördelning på $[0, a]$ är $\hat{a} = \max(X_i)$
- Är denna vvr? Alltså är $E(\hat{a}) = a$?
- För att avgöra detta måste vi ta reda på tätheten för $\hat{a}$

Egenskaper hos ML-skattaren för övre gränsen i likformig fördelning

- Vi har för $0 \leq x \leq a$ att

Om den största är mindre än x så är alla mindre än x (och vice versa)

$$\begin{aligned} P(\hat{a} \leq x) &= P(\max(X_i) \leq x) = P(X_1 \leq x, \dots, X_n \leq x) \\ &= (P(X_i \leq x))^n = \left(\frac{x}{a}\right)^n \end{aligned}$$

Oberoende

- Så tätheten för $\hat{a}$ är

$$\frac{d}{dx} \left(\frac{x}{a}\right)^n = n \frac{x^{n-1}}{a^n}$$

Egenskaper hos ML-skattaren för övre gränsen i likformig fördelning

- Det följer att (verifiera sista olikheten)

$$E(\hat{a}) = \int_0^a x n \frac{x^{n-1}}{a^n} dx = \frac{n}{n+1} a$$

- Alltså är ML-skattaren inte vvr eftersom

$$E(\hat{a}) - a = \frac{n}{n+1} a - a = -\frac{a}{n+1}$$

- Men den är asymptotiskt vvr, eller hur?

Egenskaper hos ML-skattaren för övre gränsen i likformig fördelning

- Andramomentet blir (verifiera)

$$E(\hat{a}^2) = \int_0^a x^2 n \frac{x^{n-1}}{a^n} dx = \frac{n}{n+2} a^2$$

- Så variansen är (verifiera)

$$V(\hat{a}) = \frac{n}{n+2} a^2 - \left(\frac{n}{n+1} a \right)^2 = \frac{n}{(n+2)(n+1)^2} a^2$$

Egenskaper hos ML-skattaren för övre gränsen i likformig fördelning

- Det följer att MSE är (minns; varians plus kvadrerad bias)

$$\frac{n}{(n+2)(n+1)^2} a^2 + \left(-\frac{a}{n+1} \right)^2$$
$$= \frac{2}{(n+2)(n+1)} a^2$$

Momentskattaren för övre gränsen i likformig fördelning på $[0, a]$

- Vi vet att om X är (kontinuerligt) likformigt fördelad på $[0, a]$ så är $E(X) = \frac{a}{2}$
- Det följer att momentskattaren för a är $2\bar{X}$
- Momentskattaren är alltså vvr (eller hur?)
- Detta talar till momentskattarens fördel

Momentskattarens varians

- Vi vet att

$$V(2\bar{X}) = 4V(\bar{X}) = \frac{4}{n} V(X_i) = \frac{4}{n} \frac{a^2}{12} = \frac{a^2}{3n}$$

- Eftersom momentskattaren är vvr så är MSE och varians lika

Relativ effektivitet mellan ML- och momentskattaren

- Vi har att kvoten av MSE för ML-skattaren och MSE för momentskattaren blir

$$\frac{\frac{2}{(n+2)(n+1)} a^2}{\frac{a^2}{3n}} = \frac{6n}{(n+2)(n+1)}$$

- Detta betyder att för stickprovsstorlekar större än 2 så har ML-skattaren lägre MSE vilket talar till ML-skattarens fördel vid sådana stickprovsstorlekar.